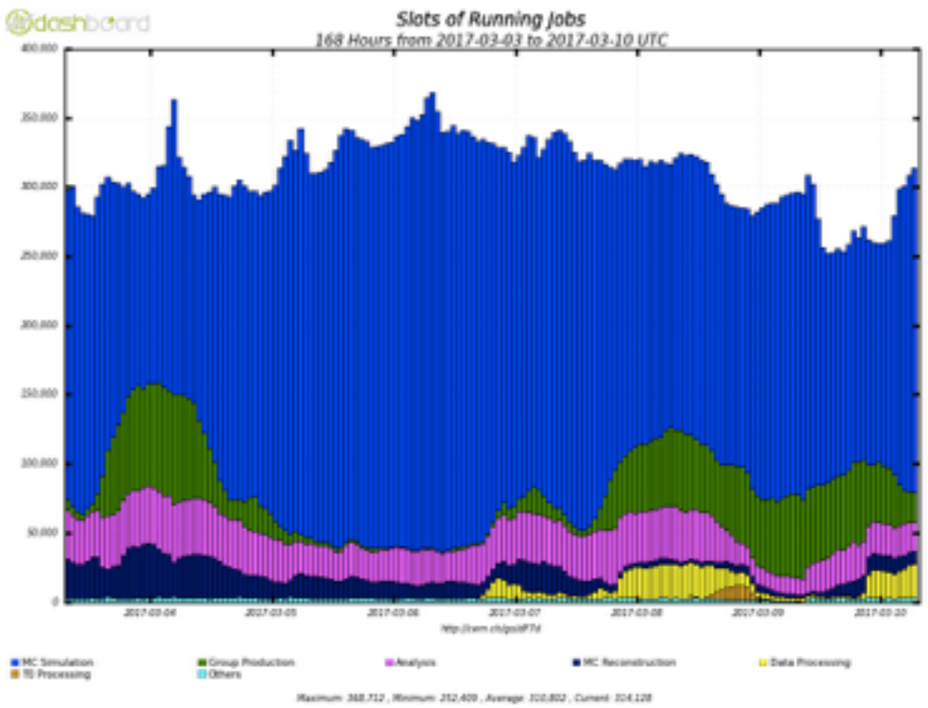
GridPP5 & Beyond: UK plans for ATLAS computing

Roger Jones

Overall Status

The general message is that we are doing very well

- Athena R21 reconstruction is ready together with the simulation software for MC16
- MC16 started, reprocessing of 2015/2016 data imminent
- MC16 and R21 will be used for the remaining of Run-2 analysis and LS-2
- It took a while to get there, need to take this into account in terms of planning for R22 (AthenaMT)



We are running 300k+ jobs/day steadily which is 20% more than we are used to (and 100% above our pledge)

Atlas current hot topics cf computing

- Containers have arrived, high interest in software and ADC, beginning support discussion with sites
  - Singularity has attractions; docker is widely adopted
  - Security concerns with private containers not an issue with singularity. (to be followed up and discussed wider)
  - Want to settle on workflows where docker or singularity images can be managed in an automated way
- The elastic search fuelled monitoring/analytics revolution continues to our advantage
- A fire has been lit under the development of the fast simulation and fast chain
- All of the technical measures to ameliorate high computing needs that ATLAS documented to LHCC have been prominent in recent activity
  - G4 sim optimisation, fast chain, overlay, pre-mixing, opportunism, workflow streamlining, throughput studies, new approaches and workflows to economise storage

Resources

We submitted our request for 2018 resources. We will discuss with the referees this week. The document is accessible to all ATLAS members. https://twiki.cern.ch/twiki/bin/view/AtlasComputing/ComputingModel

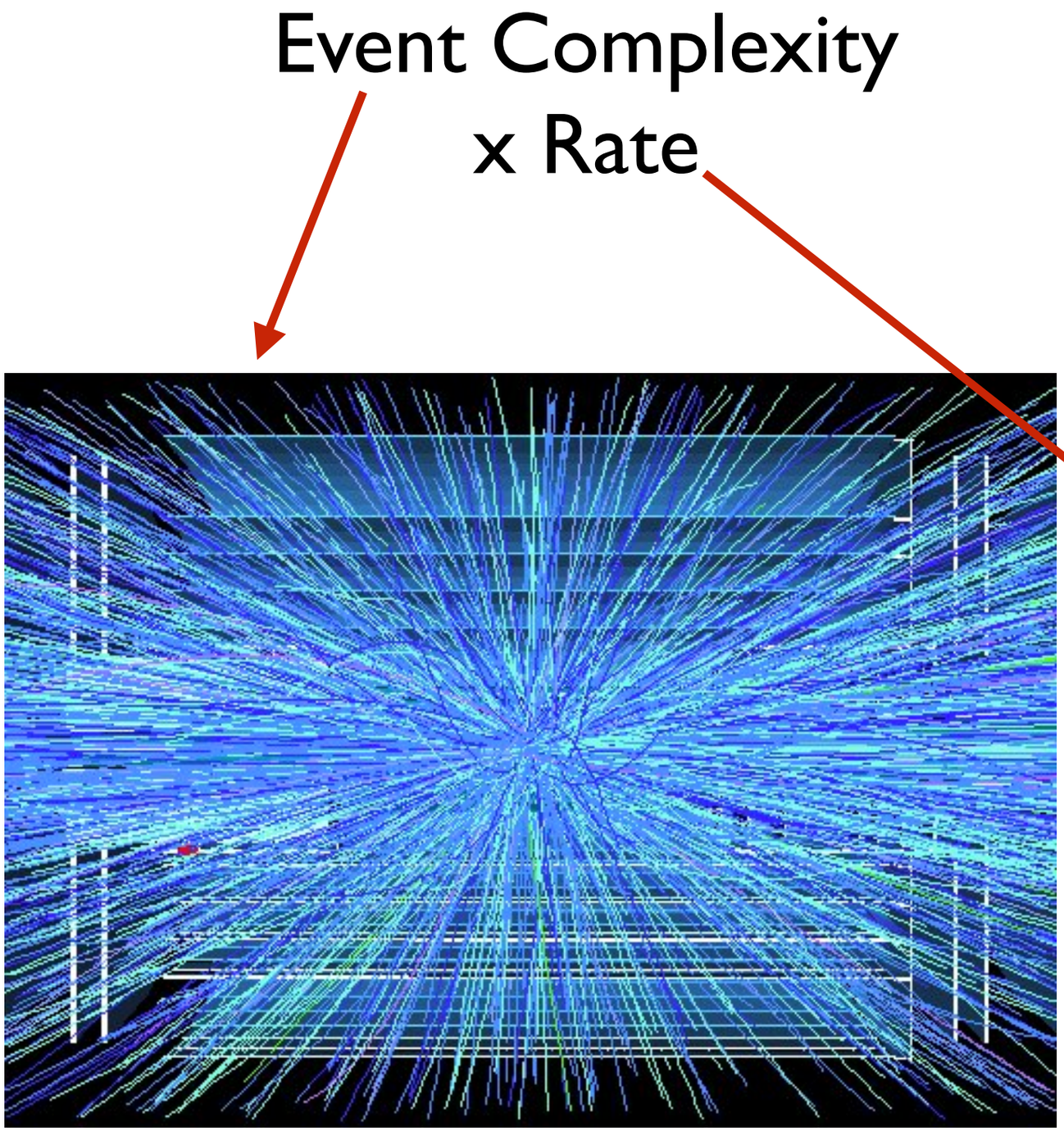
We also submitted a document to the LHCC where we examine possible mitigation options in case of Computing Resources shortage and evaluate the impact on physics. We really encourage everyone to read this: https://cds.cern.ch/record/2244528

The bottom line is we are very tight in computing resources for 2017 and 2018 and we have negligible contingency

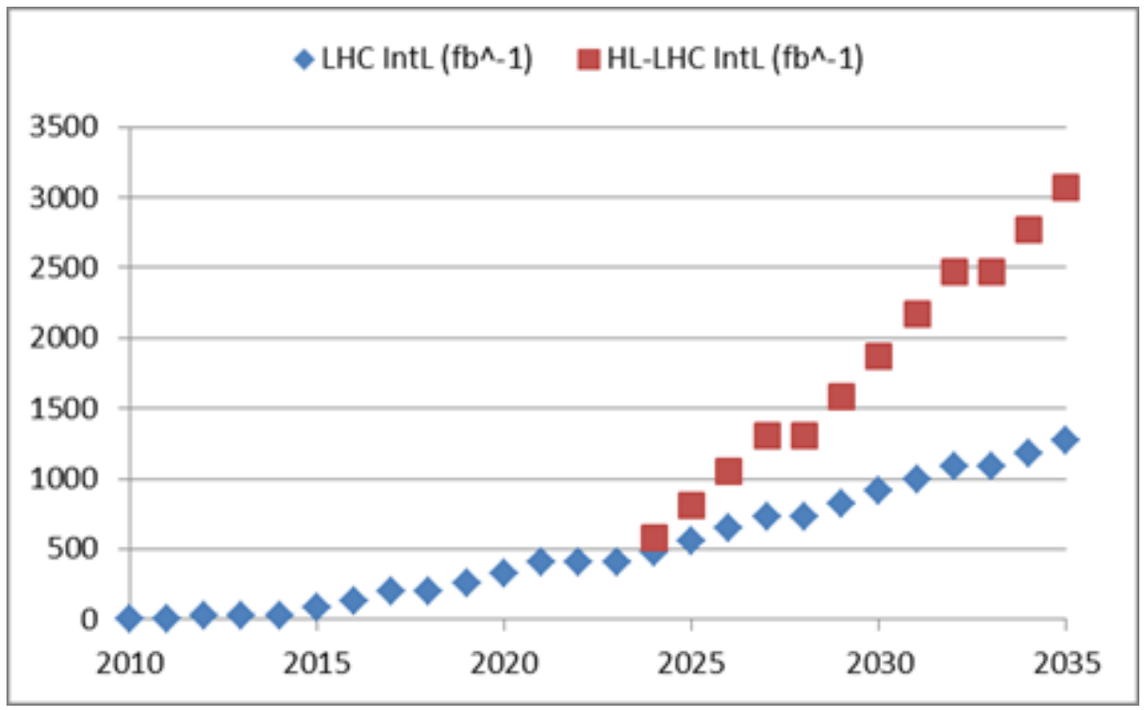
It is our job to make sure we efficiently use those resources so that physics is not impacted

	2016 pledges	2017 Req @ Oct2016 RRB	2017 Pledges	2018 Req @ Oct2016 RRB	2018 Req @ Apr2017 RRB	Balance 2018 wrt 2017 request	Balance 2018 wrt 2017 pledges
T0 CPU	257	404	404	411	411	2%	2%
T1 CPU	571	921	808	949	949	3%	17%
T2 CPU	633	1125	982	1160	1160	3%	18%
SUM CPU	1461	2450	2194	2520	2520	3%	15%
T0 DISK	17	25	25	27	26	4%	4%
T1 DISK	52	68	69	74	72	6%	4%
T2 DISK	68	83	78	91	88	6%	13%
SUM DISK	137	176	172	192	186	6%	8%
T0 TAPE	42	77	77	105	94	22%	22%
T1 TAPE	119	188	174	211	195	4%	12%
SUM TAPE	161	265	251	316	289	9%	15%

High Luminosity LHC



- Very high pile up
- Very high trigger acceptance rates
- Very challenging computing

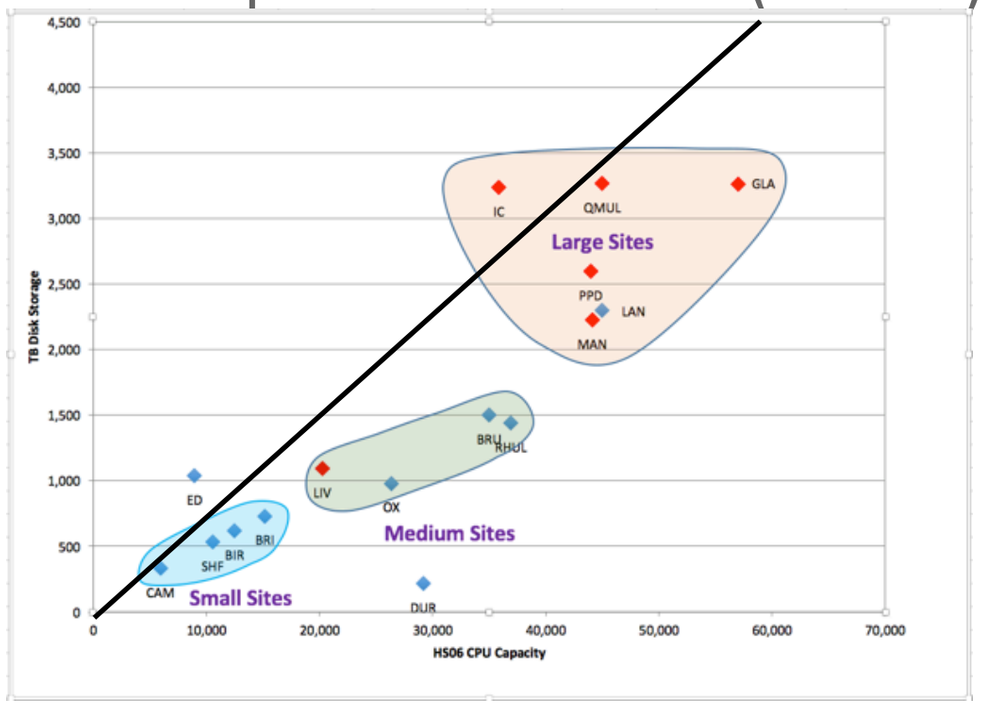


The Atlas needs

- Atlas computing is already resource limited in Run 2. This will continue in Run 3 and be considerably worse in Run 4 and beyond
  - In Run 2 and 3, a major limitation is storage.
    - CPU is useful, & under pledge; largely used for simulation, which adds to the storage problem.
    - We got ‘opportunistic’ CPU (so far) - but far less ‘opportunistic’ disk
  - We are attempting to use more fast simulation going directly to compressed formats
    - But fast simulation (& reproducibility requirements) could lead to even more pressure on storage
  - We have moved to the XAOD/train model to make every format version more useful
  - We have introduced lifetime models and popularity based replication

Atlas in the UK

- The UK is distinctive in Atlas in the number and size of sites
  - Other regions have generally invested in fewer large sites
  - The UK has leveraged local funding in many cases
    - However, the balance it has produced is heavy on CPU
    - It will need to rebalance
- We also have a lot of skilled and welcome effort (more than most regions)
  - But this is under pressure from an (at best) flat funding scenario



Atlas in the UK

- The UK is distinctive in Atlas in the number and size of sites
  - Other regions have generally invested in fewer large sites
  - The UK has leveraged local funding in many cases
    - However, the balance it has produced is heavy on CPU
    - It will need to rebalance
- We also have a lot of skilled and welcome effort (more than most regions)
  - But this is under pressure from an (at best) flat funding scenario
- We asked computing co-ordination how they would recommend we address this
  - They envisage a 3 Tier 'Tier-2+' model
  - For the UK this implies 4 full power sites (plus the Tier 1), having storage and CPUs
  - They will process and store any workflow

Non-Nucleus sites

- Medium sized sites
  - These would have significant disk, but w/o requiring the quality/service that would allow storing primaries, serving other sites for input.
  - This storage should be seen as large cache
  - The site should be able to process even in the case this cache is broken
  - It will need to rebalance
- These sites can run non trivial workflows, such as reconstruction and reprocessing, but not the most I/O demanding tasks like analysis (both trains and chaotic)

Non-Nucleus sites

- Small sized sites
  - These would have practically no disk.
  - This storage should be seen as large cache
    - This could mean no disk at all (so reading directly from some other storage)
    - Or it could mean a small disk cache allowing them to stage in and stage out
    - At a minimum - it should at least stage in inputs for simulation
- These sites will mainly run simulation etc

Cache sizes

- For both small and medium sites the cache could be "managed"
  - something like scratch disk or proddisk, i.e. a Rucio endpoint
- Or unmanaged (like a HTTP or xrootd cache).
- The choice will depend on local resources and other need
  - Some sites will deliver managed storage, for example if they will install managed storage also for the local groups and therefor it is little cost for them to use it also for processing)
- The absolute minimum disk for a small site would be 10TB; a rule of thumb is 10TB per 1000 cores.
- Medium size sites will need 100-200TB
  - The scaling is weak, ~20TB/1000 cores

Workflow

- Avoid direct copy to WN scratch from SE
  - ➡ Too stressful on the SE
- Remote IO for analysis or derivations
  - ➡ Marginal, still likely to overload the SE
- Option 1: requires fast cache disks (SSDs)
  - ➡ Caching and direct IO from cache reading through nfs, cep or rooted
- Option 2: recommended
  - ➡ Copy to scratch from cache

Institutional Commitments

- The new management in ATLAS are seeking ‘institutional commitments” for various tasks
  - This implies a long term intention to support an activity
  - There is some recognition that we do not necessarily control resource at the institute level.
- Many of the propose computing tasks are in the GridPP remit
  - I have suggested that the list of tasks should be simplified
  - They are typically “class 4”, but some are “class 3” (e.g. the production system)
- I will be entering the UK commitments in the class 4 area as taken by GridPP
  - I have suggested that the list of tasks should be simplified
  - Other computing commitments using ATLAS CG effort need to be added
  - I need to report next Tuesday, so please discuss with your PI and ask them to contact me.